



An Overview Based on the Overall Architecture of Traffic Forecasting

Lilan Peng¹ · Xuehua Liao² · Tianrui Li¹ · Xiangyu Guo¹ · Xiu Wang¹

Received: 31 August 2023 / Revised: 28 January 2024 / Accepted: 19 February 2024
© The Author(s) 2024

Abstract

With the exponential increase in the urban population, urban transportation systems are confronted with numerous challenges. Traffic congestion is common, traffic accidents happen frequently, and traffic environments are deteriorating. To alleviate these issues and improve the efficiency of urban transportation, accurate traffic forecasting is crucial. In this study, we aim to provide a comprehensive overview of the overall architecture of traffic forecasting, covering aspects such as traffic data analysis, traffic data modeling, and traffic forecasting applications. We begin by introducing existing traffic forecasting surveys and preliminaries. Next, we delve into traffic data analysis from traffic data collection, traffic data formats, and traffic data characteristics. Additionally, we summarize traffic data modeling from spatial representation, temporal representation, and spatio-temporal representation. Furthermore, we discuss the application of traffic forecasting, including traffic flow forecasting, traffic speed forecasting, traffic demand forecasting, and other hybrid traffic forecasting. To support future research in this field, we also provide information on open datasets, source resources, challenges, and potential research directions. As far as we know, this paper represents the first comprehensive survey that focuses specifically on the overall architecture of traffic forecasting.

Keywords Traffic forecasting · Graph neural network · Traffic data

1 Introduction

With the ongoing growth of urban populations, transportation problems in major cities, such as traffic congestion [1, 2] and air pollution [3, 4], are becoming increasingly severe. Intelligent transportation systems (ITS) are emerging as a crucial infrastructure in developing smart cities, and traffic forecasting plays a significant role in its advancement.

Conducting early intervention through traffic prediction can enhance the efficiency of urban transport systems, guide scheduling, and staff preallocation.

Accurate traffic prediction requires traffic big data as its support and foundation. Bello et al. [5] and Zhang et al. [6] had identified the five characteristics of big data, referred to as 5 V features, including large volume, large velocity, large variety, veracity, and value. The 5 V features highlight the notable increase in data volume. Specifically, traffic data is a sort of big data. It also has the feature of multi-source, heterogeneity, and multi-modal, which are all from different sources, such as sensors installed on the road (e.g., loop detectors), transactions in urban public transportation systems (e.g., bus or subway), smartphones equipped with GPS receivers, etc. Furthermore, traffic big data exhibits various forms of data representation, encompassing text, video, numbers, and symbols. Therefore, how to perceive, collect, and manage traffic big data is a significant issue. Besides, analyzing and exploiting the value of traffic big data presents another significant challenge.

In recent decades, numerous studies have been conducted to address the aforementioned challenges in traffic forecasting. Early, traffic forecasting was directly treated as a time

✉ Tianrui Li
trli@swjtu.edu.cn

Lilan Peng
llpeng@my.swjtu.edu.cn

Xuehua Liao
liaoxuehua@sicnu.edu.cn

Xiangyu Guo
xy_guoclay@163.com

Xiu Wang
xiuwang0214@163.com

¹ School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, Sichuan, China

² School of Computer Science, Sichuan Normal University, Chengdu 610101, Sichuan, China

series, and thus, there were many traditional time series forecasting methods applied here, e.g., Auto-Regressive Integrated Moving Average (ARIMA) [7]. Later, machine learning methods of data-driven models like Support Vector Regression (SVR) [8] were employed to handle more complex traffic data. Even though these models heavily rely on feature engineering and is often limited by the feature representation capacity. Consequently, their performance in real-world applications remains unsatisfactory due to the reliance on certain unrealistic assumptions. To overcome these limitations, many researchers have turned to deep learning techniques for modeling high-dimensional spatio-temporal data. For example, Recurrent Neural Networks (RNNs) (e.g., LSTM [9], Gated Recurrent Unit (GRU) [10]) have been utilized to capture the temporal feature of traffic data. Additionally, Convolutional Neural Networks (CNNs) have been applied to represent the spatial features of grid-based traffic data effectively. More recently, Graph Convolutional Networks (GCN) have gained popularity for learning spatial correlations in graph-based data.

There have been numerous surveys on traffic forecasting [11–20], but few studies focus on the holistic architecture of traffic forecasting. As depicted in Fig. 1, for traffic prediction, researchers generally first analyze the characteristics of traffic data. Subsequently, they consider how to design a model that can effectively capture spatial and temporal features. Finally, they devote attention to developing a loss function that enables accurate traffic prediction. In light of the above pipeline, we aim to present a comprehensive overview based on the overall architecture of traffic forecasting. To be more specific,

our initial analysis mainly focuses on traffic data collection, traffic data formats, and traffic data characteristics. Then, we conclude the methods of spatial representation modeling, temporal representation modeling, and spatio-temporal representation modeling, respectively. Lastly, we summarize the specific prediction tasks related to traffic data, such as traffic flow forecasting, traffic speed forecasting, traffic demand forecasting, and other hybrid traffic forecasting.

The contributions of this work are listed as follows:

1. We summarize and analyze traffic forecasting tasks by considering the overall architecture, including traffic data analysis, traffic data modeling, and traffic forecasting applications.
2. We review the open datasets and source resources for traffic forecasting.
3. The existing problems and challenges of traffic forecasting are presented, and the future research direction and idea are also discussed.

The rest of this paper is organized as follows: In Sect. 2, the paper introduces related surveys and traffic forecasting tasks. Sect. 3 shows traffic data analysis, including traffic data collection, traffic data formats, and traffic data characteristics. Traffic data modeling is explored in Sect. 4. The applications of traffic forecasting are introduced in Sect. 5. Section 6 provides public datasets for traffic forecasting. Sect. 7 discusses the challenges of the current traffic forecasting models and presents the future research directions. Section 8 is the conclusion of this paper.

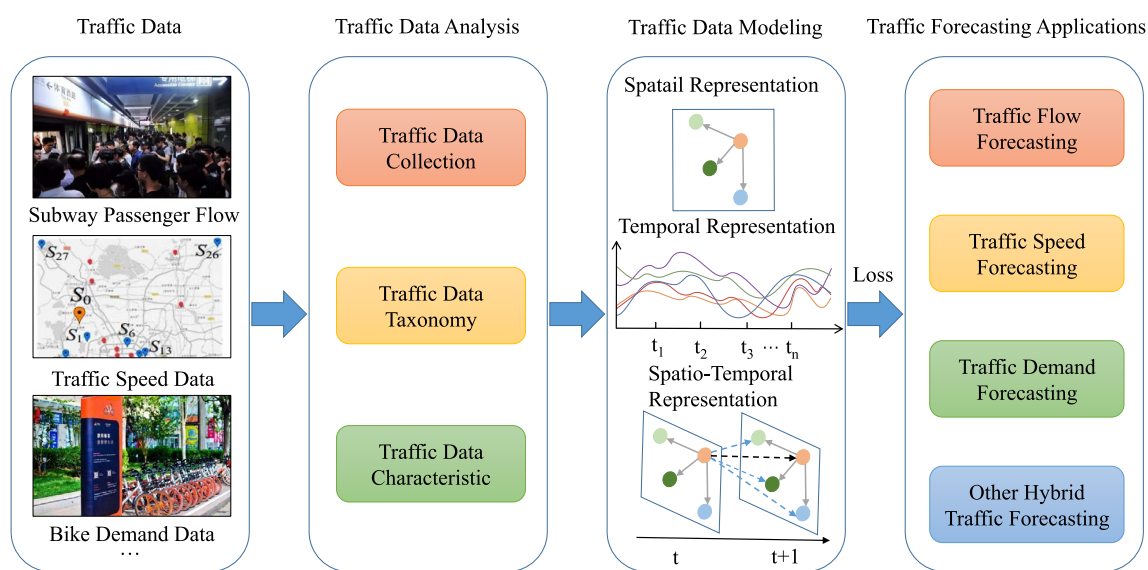


Fig. 1 The overall architecture of traffic forecasting

2 Related Work

In this section, we first make related surveys about traffic forecasting and then discuss the distinctions between existing surveys and our work. Additionally, we provide a preliminary about traffic forecasting, categorizing it into single-step and multi-step forecasting tasks.

2.1 Related Survey

The existing traffic forecasting surveys can be divided into three categories: deep learning-based, graph neural network-based, and task-based reviews. The deep learning-based short-term traffic forecasting was roughly classified into five generations in [15]. Three deep learning model categories for traffic forecasting were highlighted, including grid-based, graph-based, and multivariate time series models in [14]. The data preparation process methods and five traffic forecasting models were concluded in [16]. This survey covered spatio-temporal datasets collection technology, urban map division, and other various traffic data issues (e.g., data missing, data imbalance, and data uncertainty) as three traffic data preparation process methods. Moreover, it provided five traffic forecasting methods: statistics-based methods, machine-based methods, deep learning-based methods, reinforcement learning-based methods, and transfer learning-based methods. Yin et al. [17] reviewed traditional and deep learning methods for traffic forecasting. The traffic prediction techniques were categorized into four groups in [18], including machine learning, computational intelligence, deep learning, and hybrid methods. Traffic volume prediction and speed prediction problems were discussed in [19]. The models were only divided into statistical models, machine learning-based methods, and graph neural networks-based methods.

Due to the spatial graph structure being non-Euclidean, graph neural networks and their variants have emerged to represent the spatio-temporal relation in traffic prediction. Recently, many surveys about graph neural networks for traffic prediction have been conducted. The spatio-temporal graph neural network models were divided into graph convolutional recurrent neural network, fully graph convolutional network, graph multi-attention network, and self-learning graph structure in [11]. A GNN-based traffic forecasting survey was given in [12], which discussed diverse traffic forecasting applications, such as traffic flow forecasting, traffic speed forecasting, traffic demand forecasting, and other hybrids. Jiang et al. [13] summarized open data and big data tools applied in traffic estimation and prediction, provided different data types that are utilized for traffic estimation and forecasting tasks, and

opened a new GPS trajectory dataset for further research. Jin et al. [20] introduced spatio-temporal graph data generation methods and the deep-learning architectures in spatio-temporal graph neural networks (STGNN). It also thoroughly analyzed existing STGNN methods for temporal learning, spatial learning, and spatio-temporal fusion. Further, it examined some recently emerging approaches that integrated STGNN with other advanced learning frameworks. It mainly focused on STGNNs in various fields, including public safety, healthcare, transportation, environment, climate, and others.

In conclusion, as is mentioned in Table 1, these existing surveys mainly focus on methods or task classification without delving into the overall architecture of traffic forecasting. The differences between our study and existing surveys are that we present the overall architecture of traffic forecasting, including traffic data analysis, traffic data modeling, and traffic forecasting applications.

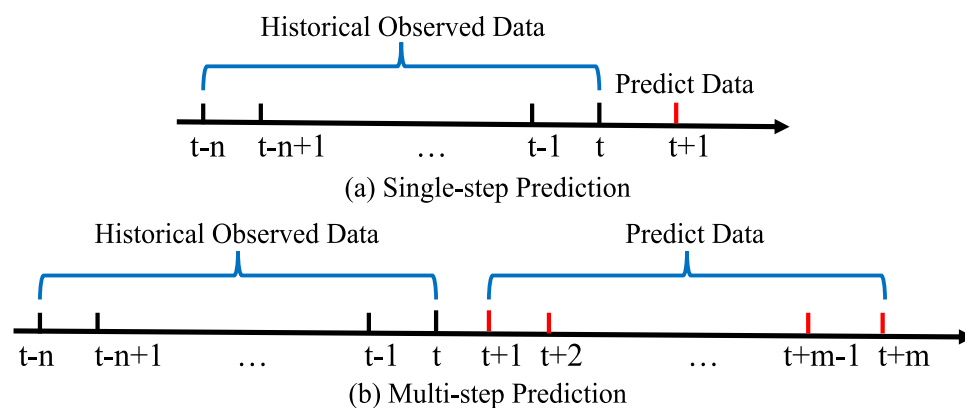
2.2 Traffic Forecasting

Traffic forecasting is an essential component of intelligent transportation systems, which generally predict the future traffic state (e.g., traffic speed, traffic flow, traffic demand, etc.) at a particular time interval when given historical observed data and other related external data. Traffic forecasting tasks are usually classified into short-term forecasting [21, 22] and long-term forecasting [23, 24] based on prediction time interval length. The time interval of short-term traffic forecasting is generally 5 min or 15 min, and the time interval of long-term traffic forecasting is 1 h or even more. In addition, according to the number of the prediction time interval, traffic prediction tasks can be divided into single-step traffic prediction [21, 22] and multi-step traffic

Table 1 Comparison between this and other related surveys

Survey	Deep learning-based	GNN-based	Task-based	Overall architecture-based
Bui et al. [11]	✓	✓	✗	✗
Jiang et al. [12]	✓	✓	✓	✗
Jiang et al. [13]	✓	✓	✓	✗
Jiang et al. [14]	✓	✓	✗	✗
Lee et al. [15]	✓	✗	✗	✗
Xie et al. [16]	✓	✓	✗	✗
Yin et al. [17]	✓	✗	✗	✗
George et al. [18]	✓	✗	✗	✗
Boukerche et al. [19]	✓	✓	✓	✗
Jin et al. [20]	✓	✓	✓	✗
This survey	✓	✓	✓	✓

Fig. 2 the example of single-step prediction and multi-step prediction



prediction tasks [23–25], which is shown in Fig. 2. We introduce the concepts of single-step traffic prediction in Definition 1 and the multi-step traffic prediction problem addressed in traffic prediction tasks in Definition 2.

Definition 1 (*Single-step Traffic Prediction*) Given historical observed traffic data of n time steps $X = (x_{t-n}, x_{t-n+1}, \dots, x_t)$, single-step traffic prediction problem aims to predict traffic data in the next time interval $t + 1$.

Definition 2 (*Multi-step Traffic Prediction*) Given observations of historical time steps $X = (x_{t-n}, x_{t-n+1}, \dots, x_t)$, multi-step traffic prediction problem targets to predict traffic data of the next m time steps, denoted as $Y = (x_{t+1}, x_{t+2}, \dots, x_{t+m})$.

3 Traffic Data Analysis

Before designing traffic prediction models, researchers are usually required to obtain traffic data, investigate the feature/pattern of traffic data, and mine existing issues of traffic data. Therefore, in this section, we initially analyze traffic data collection scenes and summarize various approaches to address different data preprocessing issues. Secondly, we conclude two traffic data formats and discuss distinctions between traffic region data and point data. Finally, we present traffic data characteristics and provide some examples. As the Fig. 3 is shown.

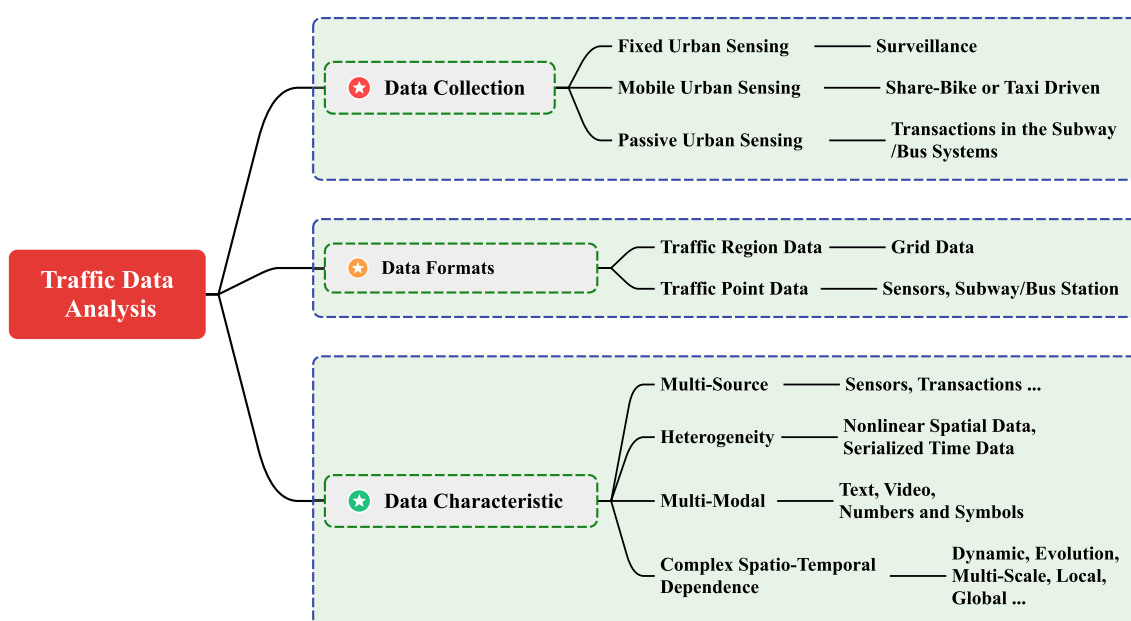


Fig. 3 Traffic data analysis for traffic forecasting

3.1 Traffic Data Collection

Many sensors have been deployed in the real world. Traffic data is mainly collected by urban sensing or automatic fare collection. Traffic data collection is mainly derived from fixed urban sensing, mobile urban sensing, and passive urban sensing. Firstly, surveillance installed on the road collects traffic speed and traffic flow data, which is a kind of fixed urban sensing. Secondly, share-bike or taxi-driven vehicles in the city collect traffic speed and traffic flow data, which is a kind of fixed urban sensing. Moreover, transactions in urban transportation systems and smartphones equipped with GPS receivers are passive urban sensing.

3.1.1 Traffic Data Pre-processing

The collected traffic data face biased samples, data sparsity, and data missing challenges. To deal with these problems, researchers developed many models for data pre-processing [26].

For biased samples of traffic data, Zhou et al. [27] designed a causal spatio-temporal graph learning framework (CauSTG) to achieve invariance in spatio-temporal data. For data sparsity, Liu et al. [28] proposed contrastive learning methods that introduced four data augmentation methods for spatio-temporal graph prediction. For data missing, main works contain traditional machine learning methods (e.g., Principal Component Analysis (PCA) [29], Matrix or Tensor Factorization (MF or TF) [30]) and deep learning methods [31, 32]. On the one hand, some scholars employ RNN and its variants to cope with traffic missing data imputation. For instance, kong et al. [31] developed a dynamic graph convolutional recurrent imputation network (DGCRIN) to impute missing traffic data. On the other hand, the adversarial learning method is integrated to traffic missing data imputation. Yuan et al. [32] proposed generative loss and center loss to minimize reconstructed errors of imputed entries and ensure

each imputed entry and its neighbors conform to local spatio-temporal distribution.

3.2 Traffic Data Formats

Based on data formats of traffic data, which are categorized into traffic region data and traffic point data, as is shown in Fig. 4. Traffic region data is like the grid data [33]. Traffic point data can include sensors and subway/bus stations [19].

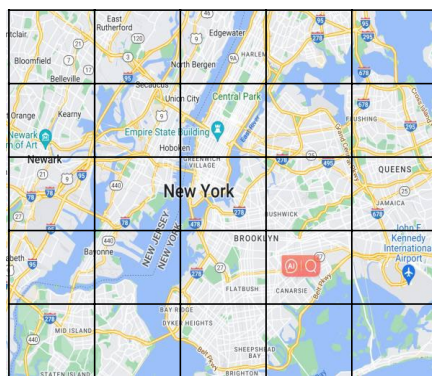
The difference between traffic region data and traffic point data is that they are in different spatial dimensions. The former is in a regular grid form, while the latter is irregular. Therefore, it can be considered that traffic region data is a special form of traffic point data when it is regularly distributed in spatial dimensions. That is, traffic point data is a more general representation than traffic region data.

To sum up, among the two typical traffic data, traffic region data is the more special type of traffic data because of its arrangement rules in time and space dimensions. Traffic point data is a kind of spatio-temporal data that is irregularly arranged in spatial dimensions, and it is a more general representation than traffic region data. Therefore, the relationship between traffic region data and traffic point data is from special to general. Thus, this paper adopts the idea of 'from special to general' and gradually summarizes the related work of two kinds of traffic data.

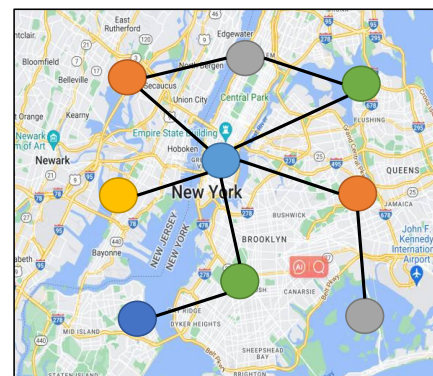
3.3 Traffic Data Characteristic

Based on the characteristics of traffic data, traffic data can be categorized into multi-source data, heterogeneity, multi-modal data, and complex spatio-temporal dependence. Researchers primarily utilize traffic big data from various sources, including sensors like loop detectors deployed in the city, transactions in urban public transportation systems, and smartphones equipped with GPS receivers. The structure of traffic big data is diverse, with heterogeneous data such as non-linear spatial data and

Fig. 4 The example of traffic data formats



Traffic Region Data



Traffic Point Data

different traffic patterns. Non-linear spatial data refers to station relations that are not regular and can be located anywhere, similar to a graph. Different traffic patterns indicate that various stations exhibit distinct travel patterns at different times. For instance, residential areas may have a strong correlation with business areas in the morning but correlate with shopping malls in the evening due to people returning home after dinner. Additionally, traffic big data encompasses various data types, such as text, video, numbers, and symbols [34]. Traffic data represents typical spatio-temporal data constantly changing with time and space, resulting in multi-scale temporal relationships and dynamic, evolutive spatial relationships. Among them, temporal correlation mainly refers to the changes in traffic status over time, resulting in data showing a certain degree of proximity, periodicity, and trend. The spatial correlation is manifested as the mutual influence of traffic conditions between various nodes of the urban road network. Besides, the functional semantics of a city can also affect its spatial relationships, resulting in nodes at longer distances exhibiting similar spatial patterns due to the same urban functions. The spatial relation also keeps the local and global correlation, such as road and region.

In addition, traffic data is also susceptible to external factors such as weather, emergencies, holidays, etc., further increasing the complexity of traffic data. To sum up, how to design an effective traffic model to obtain spatio-temporal correlations after traffic data analysis is the main problem for traffic data mining.

4 Traffic Data Modeling

Traffic data is a kind of spatio-temporal data contains spatial and temporal dimensions. Thus, we categorize traffic data modeling into three groups: spatial representation, temporal representation, and spatio-temporal representation. The spatial representation generally deals with grid data or graph data. Scholars usually employ convolution neural networks to learn the feature of grid data and learn complex and dynamic spatial dependencies of graph data through graph neural networks or their variants [35–42]. The temporal representation treats time as sequence data, researchers usually utilize RNN[43], TCN[44], Causal TCN or their variants [45, 46]. The spatio-temporal representation means that models can simultaneously capture spatial and temporal features simultaneously, such as STSGCN [47] and STJGCN [48]. Finally, we also discuss research about the combination of spatio-temporal with other promising methods, such as Meta-learning[49], ODE[23], Self-supervised learning [28], Continue learning [50] and so on.

4.1 Spatial Representation

In this section, we summarize methods of spatial representation from two parts, namely, spatial structure and spatial model.

4.1.1 Spatial Structure

The spatial structure of traffic region data is regular and a kind of grid data. For traffic point data, spatial relationships are irregular and dynamic in different time dimensions, exhibited through graph structure. The spatial information incorporates some latent semantic features except for inherent geographic features. Thus, many researchers mainly focus on traffic graph construction. We summarize traffic graph as static graphs[35–39], virtual graphs [35, 39–42, 51], hierarchical graph [53, 54, 57], or dynamic graphs [55, 56, 58–61], which are discussed in this survey and the example is shown in Fig. 5.

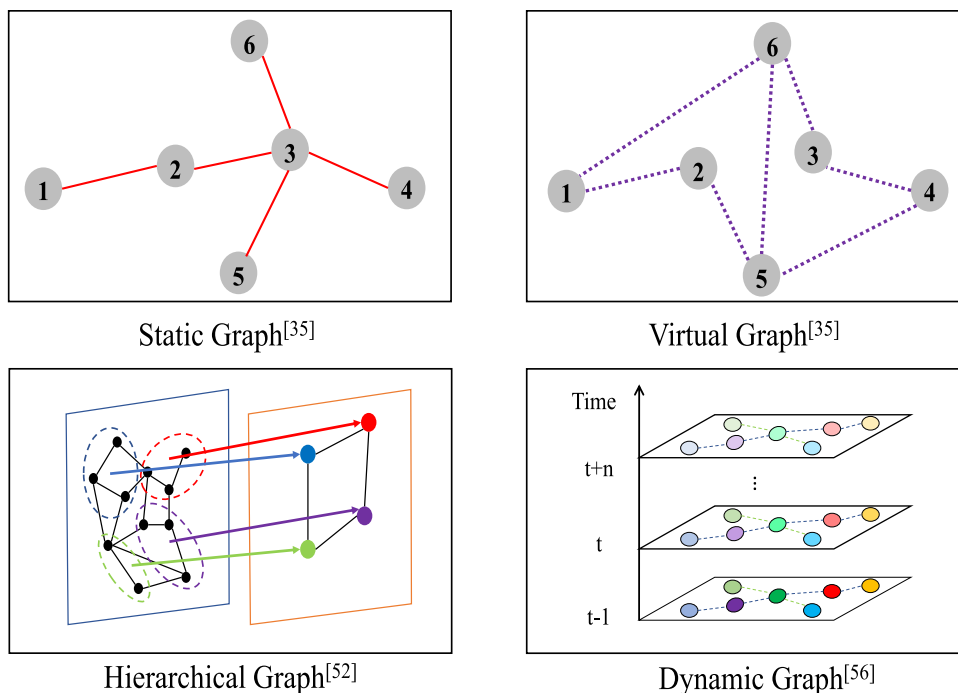
The physical graph, adjacency graph, and neighborhood graph are static graphs that can capture station geographic information. Static graphs are constructed from real-world transportation systems, such as road networks and urban public transportation systems. The functional similarity graph (e.g., POI), temporal pattern similarity graph (e.g., Dynamic Time Warping, DTW), distance graph, origin–destination graph, and heuristic graph are virtual graphs, which mainly mine the similarity of different nodes or the implicit nodes relation from various aspects. Hierarchical graphs can construct the natural hierarchical structure of traffic systems and reflect the interaction between micro and macro layers (e.g., road segments and regions). Dynamic graphs are primarily generated from data to tackle the uncertainty of spatial dependence, where node connections evolve with different temporal dimensions.

We make detailed summaries about diverse traffic graph definitions and provide how to construct the various traffic graphs in the existing works. We also provide references for each type of traffic graph in Table 2.

Table 2 The summary of traffic graph

No	Graph types	Reference
1	Static graph	[35–38, 41, 42]
2	Distance graph	[37, 39, 40]
3	Functional similarity graph	[35, 36, 39, 42, 51]
4	Temporal pattern similarity graph	[39, 52]
5	Origin–destination graph	[35–37, 42]
6	Heuristic graph	[39]
7	Hierarchical graph	[53, 54]
8	Dynamic graph	[39, 55, 56]

Fig. 5 The example of traffic graph



1. **Static Graph:** The static graph represents geographic connections. Neighbor matrix A^t can be defined as follows:

$$A_{ij}^t = \begin{cases} 1, & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

where A_{ij}^t means an element in adjacency matrix at time t , v_i and v_j are different nodes in the graph. The different works treat the static graph as different names, such as physic graph [35], neighbourhood graph [36, 38] and adjacency graph [37, 41, 42].

2. **Distance Graph:** The traffic patterns of adjacent stations can be highly correlated. For example, residents within a region may have similar daily travel patterns. Thus, the distance graph is an important factor for traffic prediction based on the perspective of geography. The distance weight matrix D is calculated using a Gaussian Kernel [62] as follows:

$$D_{ij} = \exp\left(-\frac{DS(v_i, v_j)^2}{\sigma^2}\right) \quad (2)$$

where $DS(v_i, v_j)$ indicates the shortest travel distance between different stations v_i and v_j , σ is the standard deviation of travel distances. There are different distance graph representation types, such as travel distance graph [37] and distance graph [39, 40].

3. **Functional Similarity Graph:** Generally, locations with similar functionalities or utilities (e.g., shopping

malls, schools, parks, hospitals, etc.), have strong spatio-temporal correlations. Geng et al. [36] defined the functional similarity graph A^s by using the POI similarity. The formulation is below:

$$A_{ij}^s = \text{Sim}(P_{v_i}, P_{v_j}) \in [0, 1] \quad (3)$$

where P_{v_i}, P_{v_j} are the POI vectors of regions v_i and v_j respectively, $\text{Sim}(\cdot)$ is similarity function. Moreover, Shao et al. [39] enhanced functional similarity graph W_{ij}^F by using Pearson correlation coefficients [63] to construct the global contextual function similarity graph.

$$W_{ij}^F := \begin{cases} \frac{\sum_{k=1}^K (f_{i,k} - \bar{F}_i)(f_{j,k} - \bar{F}_j)}{\sqrt{\sum_{i=1}^K (f_{i,k} - \bar{F}_i)^2} \sqrt{\sum_{j=1}^K (f_{j,k} - \bar{F}_j)^2}} & \text{if } i \neq j, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

where K is the total number of functions, then the vector of the number of global contextual similarity functions of vertex v_i is indicated as $F_i = \{f_{i,1}, f_{i,2}, \dots, f_{i,k}, \dots, f_{i,K}\}$. The function similarity graph is designed in these works [35, 36, 39, 42, 51]

4. **Temporal Pattern Similarity Graph:** The existing works mainly utilized the DTW algorithm [52] to capture temporal similarities among traffic time series of different node pairs. The similarity score $S(i, j)$ between station i and j is as follows:

$$S(i, j) = \exp(-DTW(X^i, X^j)) \quad (5)$$

Moreover, Shao et al. [39] used Pearson correlation coefficients [63] to design temporal similarity matrix W^T , which is as follows:

$$W_{ij}^T := \begin{cases} \frac{\sum_{p=1}^P (t_{i,p} - \bar{T}_i)(t_{j,p} - \bar{T}_j)}{\sqrt{\sum_{i=1}^P (t_{i,p} - \bar{T}_i)^2} \sqrt{\sum_{j=1}^P (t_{j,p} - \bar{T}_j)^2}} & \text{if } i \neq j, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

5. **Origin–Destination Graph:** The origin–destination distribution of ridership is vital for traffic prediction. Thus, Liu et al. [35] defined a correlation ratio matrix C , the formulation is as follows:

$$C(i, j) = \frac{OD(i, j)}{\sum_{k=1}^N OD(i, k)} \quad (7)$$

where $OD(i, j)$ is the total number of passengers that travel from station j to station i . The origin–destination graph has various names, such as correlation graph [35, 42], transportation connectivity [36], population flow graph [37].

6. **Heuristic Graph:** To utilize heuristic knowledge and human insights, Shao et al. [39] defined a new graph model called the heuristic graph. The formulation is as follows:

$$W_{ij}^H := \begin{cases} \exp\left(-\frac{\|d_{ij}^H\|^2}{\sigma_H^2}\right) & \text{for } i \neq j, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

The distribution distance is calculated by the Euclidean distance $d_{ij}^H = \sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2}$. σ_H^2 is a parameter to adjust the distribution of W_{ij}^H .

7. **Hierarchical Graph:** Guo et al. [53] constructed the interaction between micro and macro layers of GCNs, which integrated the different scales of features of road segments and regions for improving traffic forecasting performance.
8. **Dynamic Graph:** Xie et al. [55] employed a dynamic graph relationship learning module to learn dynamic spatial relationships between metro stations without a predefined graph adjacency matrix.

4.1.2 Spatial Model

We summarize the existing spatial representation methods and divide them into five categories: Grid-based methods, GNN-based methods, Attention-based methods, Multi-graph fusion methods, and Graph learning-based methods. We describe the details of the spatial modeling methods for the above five categories.

1. **Grid-Based Methods:** Traffic region data and image-like data have certain similarities in the data structure, and a region can be regarded as a pixel in the image. Therefore, Convolutional Neural Networks (CNNs), which have made breakthroughs in the field of image processing, have also been applied to representation learning of traffic region data, that is grid-based spatial representation methods. For example, Zhang et al. [64] proposed an ST-ResNet model based on residual convolution unit for crowd flow forecasting. Guo et al. [65] developed a spatio-temporal three-dimensional convolution neural network (ST-3DNet) to extract spatio-temporal features from traffic grid data. Some scholars combined CNN and LSTM to present a novel convolution long and short-time memory network (ConvLSTM) [66] to process spatio-temporal grid data. In general, these methods can learn effective feature representations by capturing temporal and spatial correlation in traffic grid data due to the powerful expressive ability of CNN.
2. **GNN-Based Methods:** Traffic point data is an irregular structure in the spatial dimension. Traditional CNN models cannot be directly used for representation learning of traffic graph data. Recently, due to graph convolution networks' strong ability to represent graph structure features, there are many GNN-based methods for traffic prediction. GNNs leverage a neighborhood aggregation strategy to sample and aggregate neighboring nodes features [67–69]. DCRNN [43] viewed traffic flow patterns as a diffusion process and utilized bidirectional random walks to learn spatial dependency on a directed graph. This is the first attempt to combine GNNs with RNNs for traffic forecasting. STGCN [44] developed an integration of GNNs and CNN to capture spatio-temporal correlation for traffic forecasting, which is a typical art for GNNs and CNN-based traffic forecasting.
3. **Attention-Based Methods:** After the GNN-based works, some researchers employ attention mechanisms to capture complex and dynamic spatial and temporal correlation. For instance, GMAN [24] leveraged attention mechanisms to capture spatial and temporal relationships for traffic forecasting. ASTGCN [70] employed spatial attention, temporal attention, and graph convolution network to extract dynamic spatio-temporal dependencies for traffic prediction.
4. **Multi-Graph Fusion Methods:** As is mentioned in Sect. 4.1.1, scholars designed manifold traffic graphs to learn various spatial features. How to integrate multiple traffic graphs is crucial for accurate traffic prediction. In this survey, we classify multi-graph fusion methods into three groups: static and virtual graph fusion methods, static graph and dynamic graph fusion methods, and static, virtual, and dynamic fusion methods. Firstly, there are many static and virtual graph fusion methods

[35–38, 40]. Geng et al. [36] designed neighborhood graph, functional similarity graph, and transportation connectivity to explicitly model the complex spatial relation and leveraged the contextual gated RNN to learn the temporal relation. For the second class, Li et al. [56] proposed the static graph and dynamically generated graph fusion to model complex and dynamic spatial features and used the RNN model to capture temporal information for traffic forecasting. For the third class, Shao et al. [39] constructed a static graph and four virtual graphs, then fused the five graphs into the dynamic learned graph to capture the spatial feature.

5. **Graph Learning Methods:** The graph structure is dynamic with different time series. Thus, a lot of graph learning works emerge. We divide graph learning methods into two groups: adaptive graph learning methods and graph generation methods. For adaptive graph learning works, they adaptively learn graph node features and spatial dependencies. For example, AGCRN [71] developed two adaptive modules to automatically learn node features and spatial dependencies. Wu et al. [72] designed an adaptive matrix to extract the hidden spatial dependency for traffic forecasting, which opened up graph learning research direction and brought a significant impact in traffic forecasting domains. For graph learning methods, they dynamically generate traffic graphs in different time dimensions to capture dynamic spatial correlation. For instance, Wu et al. [45] employed node embedding and edge relation learning to represent dynamic spatial dependencies. Ye et al. [46] designed a graph learning method based on multi-scale temporal dependency to learn the evolution of spatial relations, and the proposed model could reduce graph learning parameters. Zhang et al. [73] utilized static

structure learning to generate shared spatial structure and employed dynamic structure learning to model the unique structure of each node.

As is shown in Fig. 6, scholars first divide the city into several grids and utilize deep learning-based approaches to learn grid-based spatial correlation [33]. However, deep learning-based methods can't deal with non-Euclidean spatial dependencies. Thus, the GNN-based study emerges to capture complex graph structure features, such as DCRNN [43] by combining GNN and RNN, STGCN [44] with integrating GNN and CNN for traffic prediction. Afterwards, some researchers develop attention-based methods to model spatio-temporal relations. The above works mainly concentrate on the single or static graph. They can't represent complex and implicit spatial features. Therefore, multi-graph fusion methods are employed to construct various graph structures (e.g., Virtual Graph [35], Hierarchical Graph [53]) and fuse diverse graph structures to learn spatial dependencies. Traffic patterns are not the same for different stations/nodes, and spatio-temporal dependencies are dynamic and evolve in different time series. So graph-learning traffic prediction methods are applied to achieve accurate traffic prediction through node embedding and edge relation learning [45].

4.2 Temporal Representation

In this section, we review methods of temporal representation from two parts, namely, temporal feature and temporal model.

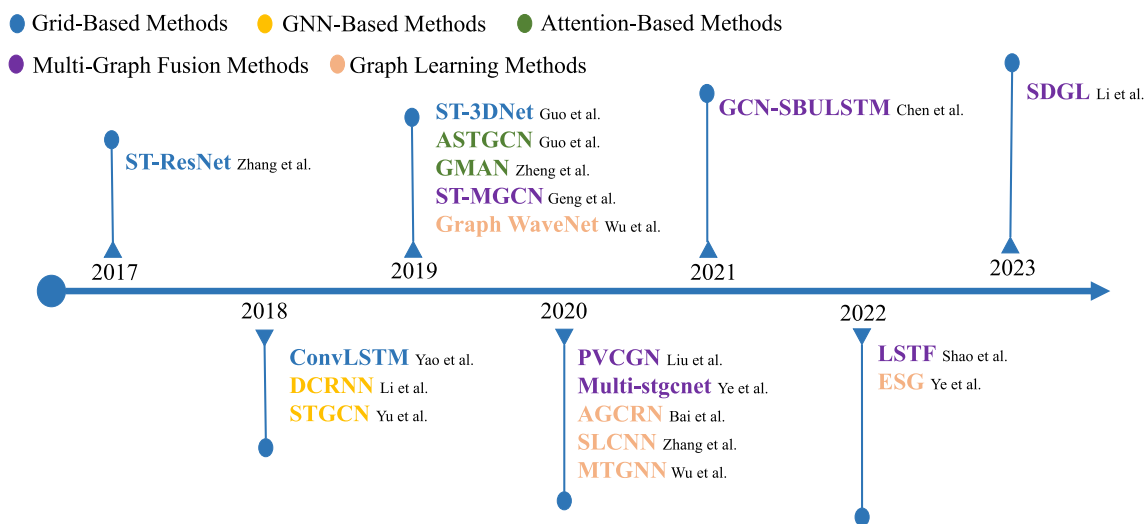


Fig. 6 A timeline of important research on spatial model for traffic prediction

4.2.1 Temporal Feature

Temporal features are categorized into closeness, periodicity, trend, daily, weekly, and holiday. The closeness is that the neighboring time series traffic trend can not change dramatically. As is shown in Fig. 7a, the traffic trend of adjacent days has only slight changes and is relatively stable. The periodicity is that traffic patterns are periodic fluctuations during different time scales. As is shown in Fig. 7b, each weekday's traffic patterns are similar and present the periodicity. The trend means traffic patterns have some fluctuation rule, such as the trend is increased during the morning peak. As shown in Fig. 7c, traffic flow on each Monday has a growth trend and a downward trend on each Saturday. The daily means the last day's traffic pattern may be similar to the next day's traffic pattern, which can be observed in Fig. 7d. The weekly relation in traffic patterns may be similar among weekdays and weekends, which can be illustrated in Fig. 7e. The traffic trend can change significantly on different holidays, as depicted in Fig. 7f.

4.2.2 Temporal Model

Traffic data is a typical time series data. Compared to time series data, 1) traffic data possesses not only temporal attributes such as closeness, periodicity, and trend but also exhibits intricate spatial relationships that are dynamic, heterogeneous, and hierarchical. Therefore, it is necessary for traffic data prediction models to incorporate spatial and temporal information learning simultaneously. It is worth mentioning that certain multivariate time series prediction

techniques that take spatial representation into account can apply to traffic prediction. For instance, MTGNN [45] and ESG [46] have shown impressive results on specific traffic datasets. However, these studies primarily focused on learning dynamic spatial features while neglecting other critical spatial attributes of traffic data, such as heterogeneity and hierarchy. Therefore, it is imperative to meticulously design specialized models that accurately represent spatial heterogeneity and hierarchy. 2) In some domains, such as electricity, incorporating spatial modeling is not required for predicting time series tasks. In fact, including spatial representation can even degrade the model's performance [74, 75]. These models are also unsuitable for traffic forecasting as they do not consider complex spatial modeling. In this section, we conclude the temporal representation methods as RNN-based, TCN-based, Causal TCN-based, and Self-attention-based methods.

1. **RNN-Based Methods:** RNN-based methods are traditional methods that regard temporal dimension as sequence data. Early RNN and its various variants were widely utilized to represent temporal features. Ye et al. [38] employed three long short-term memory network (LSTM)-based modules to learn three temporal properties (e.g., closeness, daily periodicity, weekly trend) of target stations. Zhang et al. [33] designed three residual networks to learn the closeness, period, and trend properties for traffic forecasting. Liu et al. [35] leveraged GRU to capture temporal features. Yao et al. [66] learned the spatial relations and the temporal correlations with CNN and LSTM [76] modules, respectively.

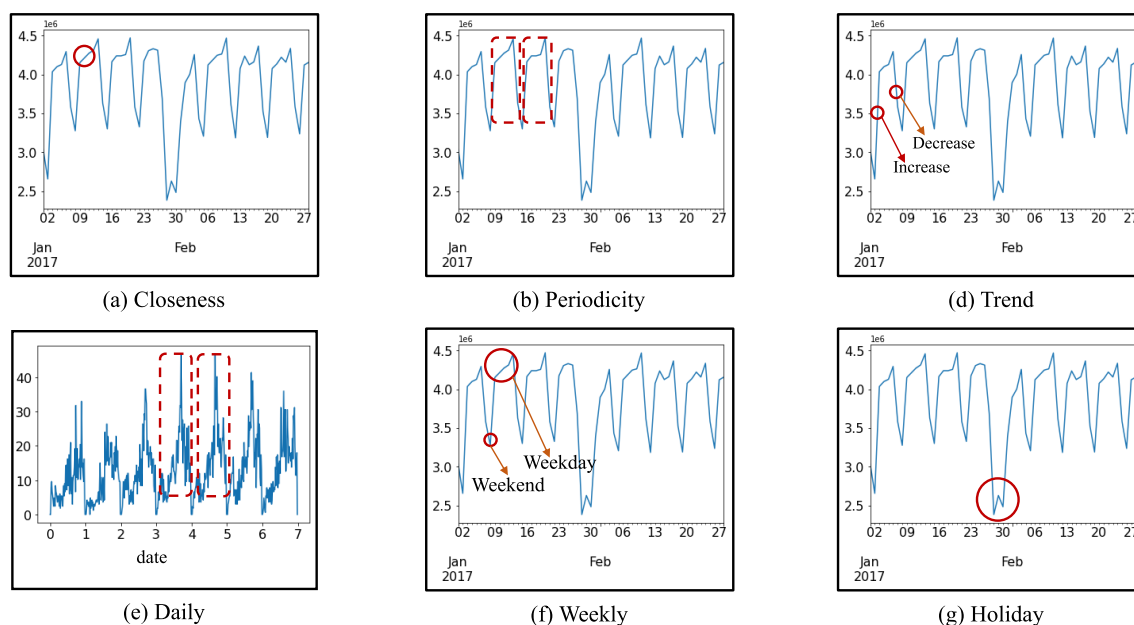


Fig. 7 The example of temporal feature

Zhao et al. [77] leveraged GRU to learn dynamic temporal dependence for traffic forecasting. RNNs have been extensively utilized in different research on traffic prediction for temporal learning. Nonetheless, these models have a notable drawback: the recurrent structures require the computation of sequences at each time step, resulting in a significant increase in computational cost and a subsequent decrease in model efficiency. Compared to RNN and its variants, Temporal Convolutional Networks (TCN) can effectively tackle this issue with their parallel 1D-CNN structures.

2. **TCN-Based Methods:** TCN-based methods employ parallel 1D-CNN structures to capture temporal features. Zhao et al. [77] proposed a temporal graph convolutional network for temporal representation, which combined GNN with gated RNN for traffic forecasting. STGCN [44] designed gated temporal convolutions to learn significant temporal features. Although TCN is an efficient parallel neural architecture for time series learning, it does not consider the temporal order of spatio-temporal graph data and multi-scale temporal features. In contrast, Causal Tcn-based methods can explicitly model the causal relation of temporal data and learn multi-scale temporal dependencies.
3. **Causal TCN-Based Methods:** Causal TCN-based methods mainly integrate the difference kernel size convolution operators to get more extensive temporal features. Wu et al. [72] applied the dilated causal convolution to model temporal features of nodes. Wu et al. [45] modified 1D convolutions with different convolutional kernels to capture temporal patterns of multiple frequencies. Ye et al. [46] employed the dilated convolution to obtain multi-scale temporal features. While Causal TCN-based methods effectively capture the causal and multi-scale relation of time series, they fall short of adequately representing long-range temporal dependencies.
4. **Self-Attention-Based Methods:** Self-attention-based methods are designed to achieve long-time series prediction, which are popular methods for long-range temporal representation. The typical model is Transformer [78]. Grigsby et al. [79] employed the transformer to learn longer temporal information. STDGRL [55] leveraged the transformer to model long-range temporal dependencies.

In conclusion, RNN-based methods [33, 35] can obtain sequence features of temporal dimensions. However, RNN-based methods rely on the output aggregated from the previous time step at each time step, making it difficult to parallelize calculations. Thus, TCN-based temporal representation methods [44, 77] are employed to learn the temporal features by parallel 1D-CNN structures. Considering the

causal affection and multi-scale attribute, causal TCN-based methods [45, 46, 72] are proposed to learn the broader range receptive field and obtain multi-scale temporal features. Furthermore, self-attention-based methods such as Transformer [78] are designed to capture long-range temporal representation.

4.3 Spatio-Temporal Representation

We analyze spatio-temporal representation methods of two types: spatio-temporal modeling simultaneously and spatio-temporal with other advanced technologies. To capture the spatio-temporal dependencies simultaneously, Song et al. [47] developed a spatio-temporal GNN model to learn localized and heterogeneous spatio-temporal dependencies synchronously. For the combination of spatio-temporal with other advanced algorithms, we divide it into four classes: Meta-learning-based methods, ODE Network-based methods, Self-supervised-based methods, and continuous Learning-based methods.

1. **Meta-Learning Based Methods:** To extract the traffic meta-knowledge and learn the diversity of spatio-temporal correlations, some scholars employ Meta-learning for traffic prediction. For instance, Fang et al. [49] designed two meta-learning-based modules to fuse multi-source external data with temporal and spatial features. ST-MetaNet [80] utilized a meta graph attention network to learn complex spatial correlations.
2. **Ordinary Differential Equation (ODE) Network-based Methods:** Graph convolutional networks generally perform better by using more stacks of layers, while the performance suffers a decrease from the depth of layers. Inspired by neural ODE network [81], Fang et al. [23] designed the tensor-based ordinary differential equation network to learn spatio-temporal relationships dynamically. To combine continuous modeling and neural ODE network, Jin et al. [82] integrated a neural ODE network with GNN for traffic forecasting.
3. **Self-Supervised Based Methods:** To solve data scarcity and heterogeneity issues, some researchers consider using self-supervised learning for traffic prediction. For example, Liu et al. [28] integrated contrastive learning with spatio-temporal GNN networks to improve the accuracy and robustness of traffic forecasting. It had a strong generalization ability and could be a plug-in for many existing spatio-temporal methods. SPGCL [83] maximized the distinctive gap between positive and negative samples and constructed an optimal graph with a self-paced strategy for traffic forecasting. Ji et al. [84] proposed a spatio-temporal contrastive learning framework, which could better represent traffic patterns and capture traffic data heterogeneity features.

4. Continual Learning Based Methods: Continual learning, also named life-long learning and incremental learning, can train the model sequentially by streaming increased data from different tasks. Chen et al. [50] introduced continual learning to obtain efficient updates and effective prediction for streaming traffic data.

In summary, it's crucial to mine the spatio-temporal data characteristics, discover traffic patterns based on traffic data, and leverage suitable and promising technology (e.g., ODE, Contrastive learning, Continual learning) to model the effective traffic prediction framework.

5 Traffic Forecasting Applications

Traffic prediction contains various applications. Here, we summarize the main applications of the existing traffic prediction work, including traffic flow forecasting, traffic speed forecasting, traffic demand forecasting, and other hybrid traffic forecasting. We also sort out and analyze the typical study for traffic forecasting in Table 3.

5.1 Traffic Flow Forecasting

Traffic flow typically connects to the number of vehicles, crowds, and passengers passing through a particular space, such as a road segment, sensor point deployed on the road, or bus/subway station in an observed time interval. Accurate traffic prediction can help to reveal real-time traffic demands and be critical for traffic management, public safety, route planning, line scheduling, and staff preallocation.

For spatial representation of traffic flow forecasting, researchers usually utilize grid-based [33], multi-graph [35] and dynamic graph methods [55]. For instance, Zhang et al. [33] partitioned cities into regular grid maps based on geographical coordinates and organized the collected traffic data as Euclidean 2D or 3D tensors, so that CNNs can be applied to extract spatial topologies. Du et al. [85] designed a hybrid multi-modal learning method to learn spatio-temporal dependencies for short-term traffic flow forecasting jointly.

These grid-based methods are suitable for predicting traffic region data but cannot model the non-linear graph structure. Therefore, several multi-graph fusion methods are proposed to learn the spatial physical and semantic information. For example, PVCN [35] could effectively capture complex ridership correlation from tailored traffic graphs. Specifically, a physical graph was directly constructed based on the real-world station topology connection. In addition, a similarity graph and a correlation graph as complementary graphs were designed to reveal the similarity and correlation of inter-station passenger flow. This paper incorporated a static graph and two virtual graphs into the graph convolution gated recurrent unit to learn the spatio-temporal relation and applied a fully connected gated recurrent unit (FC-GRU) to model the global evolution information. Finally, a Seq2Seq model with GC-GRU and FC-GRU was employed for metro ridership forecasting.

This particular method of predicting traffic by combining multiple graphs relies on predefined spatial dependencies, which are based on prior knowledge. However, the spatial relationships between traffic data constantly evolve at different time steps. Therefore, Xie et al. [55] designed a dynamic graph convolutional network to learn spatial

Table 3 The typical work for traffic forecasting

Applications	Model	Spatial representation	Temporal representation
Traffic flow forecasting	HMDLF [85]	Grid-Based	GRU
	ST-ResNet [33]	Grid-Based	LSTM
	PVCN [35]	Multi-Graph	GRU
	STDGRL [55]	Graph Learning	Transformer
	STGR [86]	GCN	LSTM
Traffic speed forecasting	DCRNN [43]	GCN	GRU
	STGCN [44]	GCN	TCN
	MTGNN [45]	Graph Learning	Causal TCN
	STSGCN [47]	GCN	TCN
	STGODE [23]	ODE	TCN
	GMAN [24]	Graph Attention	Transformer
Traffic demand forecasting	ST-MGCN [36]	Multi-Graph	GRU
	DSTGNN [58]	Dynamic Graph	Transformer
	ESG [46]	Graph Learning	Causal TCN
	CCRNN [87]	GCN	GRU
	ST-SSL [84]	Contrastive Learning	Contrastive Learning

features, leveraged transformer to obtain long-range temporal information, and employed gated fusion to combine spatial features and temporal dependencies for urban subway station passenger flow forecasting.

The typical methods for temporal representation of traffic flow forecasting are RNN and its variants for short-term traffic flow prediction [35, 85, 86], and the self-attention-based methods for long-term traffic flow prediction [55]. Li et al. [86] utilized the GNN and residual lstm for the traffic flow prediction. First, this paper calculated the correlation coefficient with min-max normalization to remove spatial heterogeneity. Then, it employed z-score transformation to eliminate daily periodicity for stronger temporal auto-correlation.

In addition, some researchers regard traffic flow prediction as a service-level/line-level task [88, 89]. For example, Luo et al. [89] proposed a spatio-temporal hashing multi-graph convolution network, where two types of sub-graphs were constructed from perspectives of physical adjacency and semantic similarity, respectively. This model explicitly captures spatio-temporal dependencies among bus stations/lines. Luo et al. [88] designed the MDL-SPFP model to jointly predict the arriving bus service flow, line-level on-board passenger flow, and line-level boarding/alighting passenger, which combines three modules, attention mechanism, residual block, and multi-scale convolution, to capture various complex non-linear spatio-temporal dependencies well.

Below, we discuss the differences among the passenger flow forecasting tasks at station, line, and bus service levels. The passenger flow forecasting at the station level is to predict the passenger flow of each station without distinguishing the passenger flow of different lines, while the passenger flow at the line level is to distinguish the passenger flow of different stations on different lines, which is a more fine-grained passenger flow forecasting. The passenger flow forecast of bus service level will further distinguish the passenger flow of different stations, lines, and vehicles. Therefore, the ridership prediction at the line/bus service level will incorporate additional prior knowledge, such as the specific line and vehicle information. Moreover, regarding spatio-temporal modeling, the inclusion of intermediate hubs and their influence will be considered.

5.2 Traffic Speed Forecasting

Traffic speed is generally based on the average speed of vehicles through certain locations, such as a sensor point in the observed time interval. This task is mainly about vehicles, researchers develop spatio-temporal modeling methods to predict the speed of vehicles, which is beneficial for travel planning.

For the spatial representation of traffic speed forecasting, sensors installed on the highway road are irregular. The

different sensors are connected through a graph structure. Thus, scholars generally use graph neural networks to capture the spatial correlation for traffic speed prediction. For the temporal representation of traffic speed forecasting, some TCN-based [44], causal TCN-based methods [45, 46] are proposed to capture different receptive temporal features. For instance, Liu et al. [43] proposed a bidirectional diffusion convolution framework to model spatial dependency and a sequence-to-sequence architecture with GRU was employed to extract temporal dependency for traffic speed forecasting. This is a pioneering work of traffic speed prediction based on the graph neural network. Afterward, researchers established different models based on the distinct characteristics of traffic speed data. To learn spatio-temporal relations synchronously, STSGCN [47] elaborately constructed a spatio-temporal synchronous modeling mechanism to learn localized spatio-temporal correlations and designed multiple modules with different periods to model spatio-temporal heterogeneity for traffic speed forecasting.

Due to shallow GNNs incapable of capturing long-range spatial correlations, Fang et al. [23] developed a tensor-based ordinary differential equation (ODE) network to model spatio-temporal dependencies for traffic speed forecasting. This work applied deeper networks to learn spatio-temporal features synchronously, which constructed a semantical adjacency matrix to obtain spatial features and elaborately designed a temporal dilated convolution structure to extract long-term temporal dependencies. To obtain the dynamic spatio-temporal dependencies, Lu et al. [90] combined the graph sequence neural network with a horizontal attention mechanism and a vertical attention mechanism to process graph sequences for traffic speed prediction.

5.3 Traffic Demand Forecasting

Traffic demand is the number of passengers with pick-up or drop-off demands, such as ride-sharing, taxi, or bike sharing, for a particular region in the observed time interval. Accurate traffic demand prediction can help to guide an efficient disposition of supplies.

For the spatial representation of traffic demand forecasting, scholars usually employ grid-based [85], multi-graph fusion [36], and dynamic graph methods [58]. For the temporal representation of traffic demand forecasting, general methods are RNN and its variants [36, 91]. Early, Du et al. [91] designed a dynamic transition CNN to obtain spatial distributions for traffic demand forecasting and to learn dynamic demand evolution. This grid-based traffic demand prediction method utilizes deep learning methods to model the spatial and temporal correlation. Then, Geng et al. [36] constructed a neighboring graph, a functional similarity graph, and a transportation connectivity graph to learn the non-Euclidean spatial dependencies. This work is

a grid-based spatial structure. The neighborhood graph was designed based on spatial proximity. The functional similarity was used by point-of-interest similarity vectors. The transportation connectivity was constructed based on the connection through motorways, highways, or public transportation. It utilized GNN to learn the three graph features fused the outputs and applied a contextual gated RNN to model temporal features for ride-hailing demand forecasting. This study employs various predefined geographic adjacency or other function graphs to represent the complex spatial semantic information, but it ignores the dynamic spatial and temporal modeling. Therefore, Huang et al. [58] developed a dynamical spatial-temporal GNN model to achieve the traffic demand prediction task.

5.4 Other Hybrid Traffic Forecasting

Besides the above three classes of traffic forecasting tasks, we summarize the other hybrid traffic forecasting tasks, such as traffic accident forecasting and delay time forecasting.

1. **Traffic Accident Forecasting:** The target of traffic accident forecasting is to predict the traffic accident numbers in the history time series. Wang et al. [51] designed a model to learn spatio-temporal dependencies from geographical and semantic perspectives for traffic accident forecasting. Zhou et al. [92] constructed a unified framework to solve urban accidents sparse issues from finer granularities and multiple steps aspects for traffic accident forecasting.
2. **Delay Time Forecasting:** The delay time of certain lines in the railway system is vital for urban public transportation management. Heglund et al. [93] utilized the sub-graph idea and constructed a graph-based formulation of a subset of the British railway network to predict the cascading delay time of the railway.

6 Public Datasets

To facilitate the participation and contribution of other researchers, we collect and organize manifold relevant open datasets about traffic forecasting tasks. A detailed description of these open datasets is provided in the following. The list of traffic forecasting datasets is shown in Table 4.

1. **TaxiBJ [33]:** This dataset contains the taxi inflow and outflow generated from more than 34,000 taxicab GPS data in Beijing from Jul 1st, 2013 to Oct 30th, Mar 1st, 2014 to Jun 30th, 2014, Mar 1st, 2015 to Jun 30th, 2015 and Nov 1st, 2015 to Apr 10th, 2016. This grid-based dataset has 32×32 grids, and the taxi flow data are counted every 30 min.
2. **SHMetro [35]:** This dataset records 811.8 million transaction records of the metro system in Shanghai. It contains 288 metro stations and 958 physical edges. The time is from Jul 1st, 2016, to Sep 30th, 2016. The time interval of each station's inflow and outflow is 15 min.
3. **HZMetro [35]** This dataset records 58.75 million transaction records of the metro system in Hangzhou. It contains 80 stations and 958 physical edges. The time slice size of each station inflow and outflow is 15 min. The time is from January 1 to January 25, 2019.
4. **SHSpeed [94]:** This dataset contains 10-minute traffic speed data for taxi trajectory data from 1 April 2015 to 30 April 2015. It includes 156 urban road segments in the central area of Shanghai.
5. **METR-LA [43]:** This dataset contains traffic speed or volume derived from the Los Angeles County road network highway. It involves 207 loop detectors. The time slice is 5 min. The time is from Mar 1st, 2012 to Jun 30th, 2012.

Table 4 The list of traffic forecasting datasets

Application	Datasets	URL
Traffic flow forecasting	TaxiBJ [33]	https://github.com/lucktroy/DeepST/tree/master/scripts/papers/AAAI17
	SHMetro [35]	https://github.com/ivechan/PVCGN
	HZMetro [35]	https://github.com/ivechan/PVCGN
Traffic speed forecasting	SHSpeed [94]	https://github.com/xxArbiter/grnn
	METR-LA [43]	https://github.com/liyaguang/DCRNN
	LargeST [96]	https://github.com/liuxu77/LargeST
	PeMS	http://pems.dot.ca.gov/
Traffic demand forecasting	Bike-NYC [91]	https://www.citibikenyc.com/system-data
	Taxi-NYC [91]	https://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
Other traffic forecasting	T-drive [95]	https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/

6. **Performance Measurement System (PeMS):** This dataset contains over 18,000 vehicle detector stations from the freeway system from all major metropolitan areas of California. The time is from 2001 to 2019. The samples are collected every 30 s and aggregated to 5 min. Each record includes a timestamp, station ID, district, freeway, direction of travel, total flow, and average speed.
7. **Bike-NYC [91]:** The dataset derives from the NYC Bike system and involves 416 stations. The period is from 1st Jul. 2013 to 31st Dec. 2016.
8. **Taxi-NYC [91]:** The taxi trip collects from both the yellow and green taxis in New York and records pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts for each trip. The time is from 2009.
9. **T-drive [95]:** This dataset contains many taxicab trajectories derived from 30,000 taxis in Beijing. The time is from Feb 1st, 2015, to Jun 2nd, 2015. It is utilized to learn smart driving directions and provide a user with the practically fastest route to a given destination at a given departure time.
10. **LargeST [96]:** This is a large-scale traffic forecasting dataset with 8,600 sensors, which consist of sensor ID, latitude, longitude, district, country, highway location, lane, type, direction. It's divided into three subsets, including Greater Los Angeles (GLA), Greater Bay Area (GBA), and San Diego (SD). This dataset provides a vital foundation for large-scale traffic prediction research.

7 Challenges and Future Directions

In this section, we discuss some general challenges for traffic forecasting tasks. Besides we also present the possible future research directions.

7.1 Challenges

We have investigated the realms of traffic data analysis, traffic data modeling, and its various applications. Despite the notable progress made in traffic prediction, several challenges remain that demand attention. We have outlined these challenges below:

1. **Heterogeneous Data:** Traffic forecasting tasks can leverage traffic data and external factors, such as weather and POI, to train the model. However, traffic data is usually confronted with missing data, sparse data, and noise. Therefore, heterogeneous traffic data fusion has been a challenging problem in the traffic domain.
2. **Multi-task Prediction:** In traffic forecasting fields, a multi-task model usually integrates all traffic information and jointly trains multiple transportation forecasting models. For instance, Li et al. [97] employed knowledge adaption to adaptive learn the knowledge from a station-intensive source to station-sparse sources for traffic demand prediction. Wang et al. [98] developed continual learning to train the model sequentially when data from various tasks obtains in a streaming manner. The main challenges are how to model the diverse data formats and the various inherent spatio-temporal correlations in different traffic forecasting tasks.
3. **Model Interpretability:** Machine learning, deep learning, and graph convolution network models have been highly successful in learning complex traffic data for traffic forecasting. However, these models are black box and end-to-end models, lacking interpretability and explainability, which have received increasing attention. While there have been studies on the explainability of Graph Neural Networks (GNN) in other domains [99–101], there is still limited research in the traffic prediction area. Therefore, a primary challenge lies in applying explainability techniques to the task of traffic forecasting [102, 103].

7.2 Future Directions

In this section, we suggest the following directions for further investigation in traffic prediction.

1. **Large-Scale Graph-Based Traffic Forecasting:** Most of the existing studies mainly consider how to deal with small-scale traffic graph data. However, many monitoring/stations in the smart city can construct a large-scale traffic graph. Therefore, it is essential to design a model to handle the traffic prediction task of a large-scale traffic graph.
2. **Self-Supervised Based Traffic Forecasting:** Graph contrastive learning has been verified effective for improving the performance of deep learning models [104], especially, data augmentation for GNNs has been identified as useful for semi-supervised node classification task [105]. However, graph contrastive learning has been explored less for spatio-temporal graphs. In particular, temporal dependencies and adaptive data augmentation are not utilized in existing graph augmentation methods. Thus, adaptive data augmentation techniques can be leveraged to identify the importance of edges/nodes to boost the performance of traffic forecasting.

3. **Knowledge Guided Traffic Forecasting:** In traffic big data, there is a lot of traffic knowledge, such as explicit or implicit traffic knowledge. The explicit traffic can be the external data, such as weather, air quality, POI, etc. The implicit traffic knowledge is defined as the hidden or latent knowledge, such as traffic potential fields [84, 106], traffic causal knowledge [107], traffic knowledge graph [108–111], etc. How to mine, represent, and integrate traffic knowledge is a hot direction for accurate traffic forecasting.
4. **Large Model Guided Traffic Forecasting:** In the urban spatio-temporal prediction domain, Huawei company proposed the Pangu large model [112] for weather forecasting, which employed a three-dimensional deep network to handle the complex weather pattern. The traffic data is multi-source (e.g., weather, accidents, etc.) and multi-modal, number, image, video, etc.). Thus, how to integrate the large model with traffic data to achieve traffic prediction is a promising research direction.

8 Conclusion

This paper presented a comprehensive review of the overall architecture of traffic forecasting. We first summarized related survey works about traffic prediction, pointed out the differences between our study and these existing surveys, and gave the traffic forecasting preliminary. Then, three critical parts of traffic forecasting were summarized: traffic data analysis, traffic data modeling, and traffic forecasting applications. We investigated traffic data from traffic data collection, traffic data formats, and traffic data characteristics. Moreover, we reviewed traffic data modeling from spatial representation, temporal representation, and spatio-temporal representation. The applications of traffic forecasting were discussed, including traffic flow forecasting, traffic speed forecasting, traffic demand, and other hybrid traffic forecasting. Furthermore, we provided the latest collection of open datasets for traffic forecasting. The challenge and future direction were also further pointed out in the following research.

Acknowledgements This research was supported by the National Natural Science Foundation of China (Nos. 62176221, 62276215, 62276218), Sichuan Science and Technology Program (No. MZGC20230073).

Author Contributions Lilan Peng: Conceptualization, Methodology, Visualization, Writing-Original draft preparation. Xuehua Liao: Conceptualization, Methodology. Tianrui Li: Writing- Reviewing, Editing, Supervision, Resources. Xiangyu Guo: Conceptualization, Methodology, Writing- Reviewing. Xiu Wang: Writing- Reviewing, Editing.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Approval and Consent to Participate The manuscript is not submitted to more than one journal for simultaneous consideration. The submitted work is original and not have been published elsewhere in any form or language (partially or in full). A single study is not be split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time (i.e. ‘salami-slicing/publishing’). Results are presented clearly, honestly, and without fabrication, falsification or inappropriate data manipulation (including image-based manipulation).

Consent for Publication All authors approved the final manuscript and the submission to this journal.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Jabbarpour MR, Zarrabi H, Khokhar RH, Shamshirband S, Choo K-KR (2018) Applications of computational intelligence in vehicle traffic congestion problem: a survey. *Soft Comput* 22(7):2299–2320
2. Arnott R, Rave T, Schöb R et al (2005) *Alleviating urban traffic congestion*. MIT Press Books, Cambridge
3. Gendron-Carrier N, Gonzalez-Navarro M, Polloni S, Turner MA (2018) *Subways and urban air pollution*. Technical report, National Bureau of Economic Research
4. Huang W, Li T, Liu J, Xie P, Du S, Teng F (2021) An overview of air quality analysis by big data techniques: Monitoring, forecasting, and traceability. *Inform Fus* 75:28–40
5. Bello-Orgaz G, Jung JJ, Camacho D (2016) Social big data: recent achievements and new challenges. *Inform Fus* 28:45–59
6. Zhang L, Xie Y, Xidao L, Zhang X (2018) Multi-source heterogeneous data fusion. In: *Proceedings of the international conference on artificial intelligence and big data*, pp 47–51
7. Ye Y, Chen L, Xue F (2019) Passenger flow prediction in bus transportation system using arima models with big data. In: *2019 international conference on cyber-enabled distributed computing and knowledge discovery (CyberC)*, pp 436–443
8. Li C, Wang X, Cheng Z, Bai Y (2020) Forecasting bus passenger flows by using a clustering-based support vector regression approach. *IEEE Access* 8:19717–19725
9. Cui Z, Ke R, Pu Z, Wang Y (2020) Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values. *Trans Res Part C: Emerg Technol* 118:102674
10. Agarap AFM (2018) A neural network architecture combining gated recurrent unit (gru) and support vector machine (svm) for

- intrusion detection in network traffic data. In: Proceedings of the international conference on machine learning and computing, pp 26–30
11. Bui K-HN, Cho J, Yi H (2022) Spatial-temporal graph neural network for traffic forecasting: an overview and open research issues. *Appl Intell* 52(3):2763–2774
 12. Jiang W, Luo J (2022) Graph neural network for traffic forecasting: a survey. *Expert Syst Appl* 207:117921
 13. Jiang W, Luo J (2022) Big data for traffic estimation and prediction: a survey of data and tools. *Appl Syst Innovat* 5(1):23
 14. Jiang R, Yin D, Wang Z, Wang Y, Deng J, Liu H, Cai Z, Deng J, Song X, Shibasaki R (2021) DI-traffic: Survey and benchmark of deep learning models for urban traffic prediction. In: Proceedings of the ACM international conference on information & knowledge management, pp 4515–4525
 15. Lee K, Eo M, Jung E, Yoon Y, Rhee W (2021) Short-term traffic prediction with deep neural networks: a survey. *IEEE Access* 9:54739–54756
 16. Xie P, Li T, Liu J, Du S, Yang X, Zhang J (2020) Urban flow prediction from spatiotemporal data using machine learning: a survey. *Inform Fus* 59:1–12
 17. Yin X, Wu G, Wei J, Shen Y, Qi H, Yin B (2020) A comprehensive survey on traffic prediction. *arXiv preprint arXiv:2004.08555*
 18. George S, Santra AK (2020) Traffic prediction using multifaceted techniques: a survey. *Wireless Pers Commun* 115(2):1047–1106
 19. Boukerche A, Tao Y, Sun P (2020) Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. *Comput Netw* 182:107484
 20. Jin G, Liang Y, Fang Y, Huang J, Zhang J, Zheng Y (2023) Spatio-temporal graph neural networks for predictive learning in urban computing: a survey. *arXiv preprint arXiv:2303.14483*
 21. Ou J, Sun J, Zhu Y, Jin H, Liu Y, Zhang F, Huang J, Wang X (2020) Stp-trellisnets: Spatial-temporal parallel trellisnets for metro station passenger flow prediction. In: Proceedings of the ACM international conference on information & knowledge management, pp 1185–1194
 22. Zhu S, Zhao Y, Zhang Y, Li Q, Wang W, Yang S (2020) Short-term traffic flow prediction with wavelet and multi-dimensional taylor network model. *IEEE Trans Intell Transp Syst* 22(5):3203–3208
 23. Fang Z, Long Q, Song G, Xie K (2021) Spatial-temporal graph ode networks for traffic flow forecasting. In: Proceedings of the ACM SIGKDD conference on knowledge discovery & data mining, pp 364–373
 24. Zheng C, Fan X, Wang C, Qi J (2020) Gman: A graph multi-attention network for traffic prediction. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp 1234–1241
 25. Liu D, Wang J, Shang S, Han P (2022) Msdr: Multi-step dependency relation networks for spatial temporal forecasting. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, pp 1042–1050
 26. Yi X, Zheng Y, Zhang J, Li T (2016) St-mvl: filling missing values in geo-sensory time series data. In: Proceedings of the international joint conference on artificial intelligence
 27. Zhou Z, Huang Q, Yang K, Wang K, Wang X, Zhang Y, Liang Y, Wang Y (2023) Maintaining the status quo: Capturing invariant relations for ood spatiotemporal learning, pp 3603–3614
 28. Liu X, Liang Y, Huang C, Zheng Y, Hooi B, Zimmermann R (2022) When do contrastive learning signals help spatio-temporal graph forecasting? In: Proceedings of the 30th international conference on advances in geographic information systems, pp 1–12
 29. Audigier V, Husson F, Josse J (2014) Multiple imputation for continuous variables using a bayesian principal component analysis†. *J Stat Comput Simul* 86:2140–2156
 30. Chen X, Sun L (2022) Bayesian temporal factorization for multidimensional time series prediction. *IEEE Trans Pattern Anal Mach Intell* 44(9):4659–4673
 31. Kong X, Zhou W, Shen G, Zhang W, Liu N, Yang Y (2023) Dynamic graph convolutional recurrent imputation network for spatiotemporal traffic missing data. *Knowl-Based Syst* 261:110188
 32. Yuan Y, Zhang Y, Wang B, Peng Y, Hu Y, Yin B (2022) Stgan: Spatio-temporal generative adversarial network for traffic data imputation. *IEEE Trans. Big Data* 9(1):200–211
 33. Zhang J, Zheng Y, Qi D (2017) Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the AAAI conference on artificial intelligence
 34. Liu J, Li T, Ji S, Xie P, Du S, Teng F, Zhang J (2021) Urban flow pattern mining based on multi-source heterogeneous data fusion and knowledge graph embedding. *IEEE Trans. Knowl Data Eng*
 35. Liu L, Chen J, Wu H, Zhen J, Li G, Lin L (2020) Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction. *IEEE Trans Intell Transp Syst*
 36. Geng X, Li Y, Wang L, Zhang L, Yang Q, Ye J, Liu Y (2019) Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence 33:3656–3663
 37. Chen P, Fu X, Wang X (2021) A graph convolutional stacked bidirectional unidirectional-1stm neural network for metro ridership prediction. *IEEE Trans Intell Transp Syst* 23(7):6950–6962
 38. Ye J, Zhao J, Ye K, Xu C (2020) Multi-stgcnet: A graph convolution based spatial-temporal framework for subway passenger flow forecasting. In: Proceedings of the international joint conference on neural networks, pp 1–8
 39. Shao W, Jin Z, Wang S, Kang Y, Xiao X, Menouar H, Zhang Z, Zhang J, Salim F (2022) Long-term spatio-temporal forecasting via dynamic multiple-graph attention. In: Proceedings of the Thirty-First international joint conference on artificial intelligence, pp 2225–2232
 40. Chai D, Wang L, Yang Q (2018) Bike flow prediction with multi-graph convolutional networks. In: Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems, pp 397–400
 41. Li F, Yan H, Jin G, Liu Y, Li Y, Jin D (2022) Automated spatio-temporal synchronous modeling with multiple graphs for traffic prediction. In: Proceedings of the ACM International Conference on Information & Knowledge Management, pp 1084–1093
 42. Shi H, Yao Q, Guo Q, Li Y, Zhang L, Ye J, Li Y, Liu Y (2020) Predicting origin-destination flow via multi-perspective graph convolutional network. In: Proceedings of the international conference on data engineering (ICDE), pp 1818–1821
 43. Li Y, Yu R, Shahabi C, Liu Y (2018) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: Proceedings of the 6th international conference on learning representations
 44. Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: Proceedings of the Twenty-seventh international joint conference on artificial intelligence, pp 3634–3640
 45. Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C (2020) Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery & data mining, pp 753–763
 46. Ye J, Liu Z, Du B, Sun L, Li W, Fu Y, Xiong H (2022) Learning the evolutionary and multi-scale graph structure for multivariate time series forecasting. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, pp 2296–2306

47. Song C, Lin Y, Guo S, Wan H (2020) Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp 914–921
48. Zheng C, Fan X, Pan S, Jin H, Peng Z, Wu Z, Wang C, Yu P (2023) Spatio-temporal joint graph convolutional networks for traffic forecasting. *IEEE Trans Knowl Data Eng*, pp 1–14
49. Fang S, Pan X, Xiang S, Pan C (2020) Meta-msnet: meta-learning based multi-source data fusion for traffic flow prediction. *IEEE Signal Process Lett* 28:6–10
50. Chen X, Wang J, Xie K (2021) Trafficstream: A streaming traffic flow forecasting framework based on graph neural networks and continual learning. In: Proceedings of the thirtieth international joint conference on artificial intelligence, pp 3620–3626
51. Wang B, Lin Y, Guo S, Wan H (2021) Gsnet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35, pp 4402–4409
52. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. *KDD Workshop* 10:359–370
53. Guo K, Hu Y, Sun Y, Qian S, Gao J, Yin B (2021) Hierarchical graph convolution network for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence 35:151–159
54. Wu N, Zhao XW, Wang J, Pan D (2020) Learning effective road network representation with hierarchical graph neural networks. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery & data mining, pp 6–14
55. Xie P, Ma M, Li T, Ji S, Du S, Yu Z, Zhang J (2023) Spatio-temporal dynamic graph relation learning for urban metro flow prediction. *IEEE Trans Knowl Data Eng* 35(10):9973–9984
56. Li ZL, Zhang GW, Yu J, Xu LY (2023) Dynamic graph structure learning for multivariate time series forecasting. *Pattern Recogn* 138:109423
57. Zhao J, Wang X, Shi C, Hu B, Song G, Ye Y (2021) Heterogeneous graph structure learning for graph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence 35:4697–4705
58. Huang F, Yi P, Wang J, Li M, Peng J, Xiong X (2022) A dynamical spatial-temporal graph neural network for traffic demand prediction. *Inf Sci* 594:286–304
59. Jin M, Li Y-F, Zheng Y, Yang B, Pan S (2021) Spatiotemporal representation learning on time series with dynamic graph odes
60. Li F, Feng J, Yan H, Jin G, Yang F, Sun F, Jin D, Li Y (2023) Dynamic graph convolutional recurrent network for traffic prediction: benchmark and solution. *ACM Trans Knowl Discov Data* 17(1):1–21
61. Jin G, Xi Z, Sha H, Feng Y, Huang J (2022) Deep multi-view graph-based network for citywide ride-hailing demand prediction. *Neurocomputing* 510:79–94
62. Kusano G, Hiraoka Y, Fukumizu K (2016) Persistence weighted gaussian kernel for topological data analysis. In: International conference on machine learning, pp 2004–2013
63. Zhang X, Cao R, Zhang Z, Xia Y (2020) Crowd flow forecasting with multi-graph neural networks. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp 1–7
64. Zhang J, Zheng Y, Qi D, Li R, Yi X, Li T (2018) Predicting city-wide crowd flows using deep spatio-temporal residual networks. *Artif Intell* 259:147–166
65. Guo S, Lin Y, Li S, Chen Z, Wan H (2019) Deep spatial-temporal 3d convolutional neural networks for traffic data forecasting. *IEEE Trans Intell Transp Syst* 20(10):3913–3926
66. Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, Gong P, Ye J, Li Z (2018) Deep multi-view spatial-temporal network for taxi demand prediction. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32
67. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907)
68. Hamilton W, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs. *Adv Neural Inform Process Syst*, 30
69. Long Q, Jin Y, Wu Y, Song G (2021) Theoretically improving graph neural networks via anonymous walk graph kernels. In: Proceedings of the web conference, pp 1204–1214
70. Guo S, Lin Y, Feng N, Song C, Wan H (2019) Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the AAAI conference on artificial intelligence 33:922–929
71. Bai L, Yao L, Li C, Wang X, Wang C (2020) Adaptive graph convolutional recurrent network for traffic forecasting. *Adv Neural Inf Process Syst* 33:17804–17815
72. Wu Z, Pan S, Long G, Jiang J, Zhang C (2019) Graph wavenet for deep spatial-temporal graph modeling. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence, pp 1907–1913
73. Zhang Q, Chang J, Meng G, Xiang S, Pan C (2020) Spatio-temporal graph structure learning for traffic forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp 1177–1185
74. Zeng A, Chen M, Zhang L, Xu Q (2023) Are transformers effective for time series forecasting?, vol. 37, pp 11121–11128
75. Nie Y, Nguyen NH, Sinthong P, Kalagnanam J (2023) A time series is worth 64 words: Long-term forecasting with transformers. In: Proceedings of the international conference on learning representations
76. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-c (2015) Convolutional lstm network: a machine learning approach for precipitation nowcasting. *Adv Neural Inform Process Syst*, 28
77. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2019) T-gcn: a temporal graph convolutional network for traffic prediction. *IEEE Trans Intell Transp Syst* 21(9):3848–3858
78. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Adv Neural Inform Process Syst*, 30
79. Grigsby J, Wang Z, Qi Y (2021) Long-range transformers for dynamic spatiotemporal forecasting. arXiv preprint [arXiv:2109.12218](https://arxiv.org/abs/2109.12218)
80. Pan Z, Liang Y, Wang W, Yu Y, Zheng Y, Zhang J (2019) Urban traffic prediction from spatio-temporal data using deep meta learning. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery & data mining, pp 1720–1730
81. Chen RT, Rubanova Y, Bettencourt J, Duvenaud DK (2018) Neural ordinary differential equations. *Adv Neural Inform Process Syst*, 31
82. Jin M, Zheng Y, Li Y-F, Chen S, Yang B, Pan S (2022) Multivariate time series forecasting with dynamic graph neural odes. *IEEE Trans Knowl Data Eng*
83. Li R, Zhong T, Jiang X, Trajcevski G, Wu J, Zhou F (2022) Mining spatio-temporal relations via self-paced graph contrastive learning. In: Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining, pp 936–944
84. Ji J, Wang J, Huang C, Wu J, Xu B, Wu Z, Zhang J, Zheng Y (2023) Spatio-temporal self-supervised learning for traffic flow prediction. In: Proceedings of the AAAI conference on artificial intelligence 37:4356–4364
85. Du S, Li T, Gong X, Yu Z, Horng S-J (2018) A hybrid method for traffic flow forecasting using multimodal deep learning. *Int J Comput Intell Syst*, 13
86. Li B, Li X, Zhang Z, Wu F (2019) Spatio-temporal graph routing for skeleton-based action recognition. In: Proceedings of the AAAI conference on artificial intelligence 33:8561–8568

87. Ye J, Sun L, Du B, Fu Y, Xiong H (2021) Coupled layer-wise graph convolution for transportation demand prediction. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35, pp 4617–4625
88. Luo D, Zhao D, Ke Q, You X, Liu L, Zhang D, Ma H, Zuo X (2020) Fine-grained service-level passenger flow prediction for bus transit systems based on multitask deep learning. *IEEE Trans Intell Transp Syst* 22(11):7184–7199
89. Luo D, Zhao D, Ke Q, You X, Liu L, Ma H (2021) Spatiotemporal hashing multigraph convolutional network for service-level passenger flow forecasting in bus transit systems. *IEEE Internet Things J* 9(9):6803–6815
90. Lu Z, Lv W, Xie Z, Du B, Xiong G, Sun L, Wang H (2022) Graph sequence neural network with an attention mechanism for traffic speed prediction. *ACM Trans Intell Syst Technol* 13(2):1–24
91. Du B, Hu X, Sun L, Liu J, Qiao Y, Lv W (2020) Traffic demand prediction based on dynamic transition convolutional neural network. *IEEE Trans Intell Transp Syst* 22(2):1237–1247
92. Zhou Z, Wang Y, Xie X, Chen L, Zhu C (2020) Foresee urban sparse traffic accidents: a spatiotemporal multi-granularity perspective. *IEEE Trans Knowl Data Eng*
93. Heglund JS, Taleongpong P, Hu S, Tran HT (2020) Railway delay prediction with spatial-temporal graph convolutional networks. In: Proceedings of the international conference on intelligent transportation systems, pp 1–6
94. Wang X, Chen C, Min Y, He J, Yang B, Zhang Y (2018) Efficient metropolitan traffic prediction based on graph recurrent neural network. *arXiv preprint [arXiv:1811.00740](https://arxiv.org/abs/1811.00740)*
95. Yuan J, Zheng Y, Zhang C, Xie W, Xie X, Sun G, Huang Y (2010) T-drive: driving directions based on taxi trajectories. In: Proceedings of the SIGSPATIAL international conference on advances in geographic information systems, pp 99–108
96. Liu X, Xia Y, Liang Y, Hu J, Wang Y, Bai L, Huang C, Liu Z, Hooi B, Zimmermann R (2023) Largest: A benchmark dataset for large-scale traffic forecasting. *arXiv preprint [arXiv:2306.08259](https://arxiv.org/abs/2306.08259)*
97. Li C, Bai L, Liu W, Yao L, Waller ST (2020) Knowledge adaptation for demand prediction based on multi-task memory neural network. In: Proceedings of the ACM international conference on information & knowledge management, pp 715–724
98. Wang J, Song G, Wu Y, Wang L (2020) Streaming graph neural networks via continual learning. In: Proceedings of the ACM international conference on information & knowledge management, pp 1515–1524
99. Baldassarre F, Azizpour H (2019) Explainability techniques for graph convolutional networks. *arXiv preprint [arXiv:1905.13686](https://arxiv.org/abs/1905.13686)*
100. Pope PE, Kolouri S, Rostami M, Martin CE, Hoffmann H (2019) Explainability methods for graph convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10772–10781
101. Ying Z, Bourgeois D, You J, Zitnik M, Leskovec J (2019) Gnnexplainer: Generating explanations for graph neural networks. *Adv Neural Inform Process Syst*, 32
102. Wu Y, Tan H, Qin L, Ran B, Jiang Z (2018) A hybrid deep learning based traffic flow prediction method and its understanding. *Transp Res Part C: Emerg Technol* 90:166–180
103. Barredo-Arrieta A, Laña I, Del Ser J (2019) What lies beneath: A note on the explainability of black-box machine learning models for road traffic forecasting. In: Proceedings of the intelligent transportation systems conference, pp 2232–2237
104. Zhu Y, Xu Y, Yu F, Liu Q, Wu S, Wang L (2021) Graph contrastive learning with adaptive augmentation. In: Proceedings of the web conference, pp 2069–2080
105. Zhao T, Liu Y, Neves L, Woodford O, Jiang M, Shah N (2021) Data augmentation for graph neural networks. In: Proceedings of the AAAI conference on artificial intelligence 35:11015–11023
106. Ji J, Wang J, Jiang Z, Ma J, Zhang H (2020) Interpretable spatiotemporal deep learning model for traffic flow prediction based on potential energy fields. In: 2020 IEEE international conference on data mining (ICDM), pp 1076–1081
107. Deng P, Zhao Y, Liu J, Jia X, Wang M (2023) Spatio-temporal neural structural causal models for bike flow prediction. In: Proceedings of the AAAI conference on artificial intelligence 37:4242–4249
108. Liu Y, Ding J, Fu Y, Li Y (2023) Urbankg: an urban knowledge graph system. *ACM Trans Intell Syst Technol* 14(4):1–25
109. Wang H, Yu Q, Liu Y, Jin D, Li Y (2021) Spatio-temporal urban knowledge graph enabled mobility prediction. In: Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 5(4):1–24
110. Wang S, Lv Y, Peng Y, Piao X, Zhang Y (2022) Metro traffic flow prediction via knowledge graph and spatiotemporal graph neural network. *J Adv Transp*
111. Zhu J, Han X, Deng H, Tao C, Zhao L, Wang P, Lin T, Li H (2022) Kst-gcn: A knowledge-driven spatial-temporal graph convolutional network for traffic forecasting. *IEEE Trans Intell Transp Syst* 23(9):15055–15065
112. Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q (2023) Accurate medium-range global weather forecasting with 3d neural networks. *Nature* 619:1–6